A Framework for Assessment Design in the Era of Generative AI: The Case of Take-Home Assignment in Software-related Courses

new Lyza FELIPEa\*, Quang Ngoc TRANb, Thanh Chi PHAMc, Thanh Ngoc NGUYENd & Linh Duc TRANe

a,b,c,d,e*School of Science, Engineering & Technology, RMIT University, Vietnam*

\*anna.felipe@rmit.edu.vn

**Abstract:** The use of generative AI in education has brought about various challenges, especially in the area of assessment design, particularly concerning academic integrity. Plagiarism is a significant concern as students can easily use AI-generated content to cheat in assignments, compromising the credibility and reliability of the assessment results. This paper addresses this issue by proposing a framework for designing take-home assignments in software-related courses incorporating generative AI and promoting academic integrity. The framework provides guidelines for designing assignments that prevent plagiarism and ensure that the assessment tasks align with the course learning objectives. It offers a practical and flexible approach to integrating generative AI into assessments while mitigating the risks associated with academic misconduct.

**Keywords:** AI Education, Bloom Taxonomy, Computing Education, Generative AI, Assessment Design

1. Introduction

In recent years, the field of education has witnessed a significant shift toward the integration of technology, particularly in the assessment design {Almond, 2002 #3529;Gorin, 2006 #3530;Van den Berg, 2006 #3531;Villarroel, 2018 #3532}. One area where this integration has gained significant attention is the use of generative AI in assessment tasks {Fergus, 2023 #3533;Geerling, 2023 #3534}. Generative AI refers to computer algorithms that can generate new content or responses based on a set of input data {Yeadon, 2023 #3535;Yeadon, 2023 #3535}. The integration of generative AI in assessment tasks provides various benefits such as efficient grading, increased objectivity, and reduced bias in the assessment {Benuyenah, 2023 #3536}. However, the use of generative AI in assessment tasks poses several challenges, especially in the area of academic integrity {Kooli, 2023 #3537}.

Academic integrity is a critical aspect of education that ensures that students maintain high ethical standards in their academic work {Emenike, 2023 #3538;Fergus, 2023 #3533}. Plagiarism, the act of presenting someone else's work as one's own, is a significant concern in academic integrity, as it undermines the credibility and reliability of assessment results {Bretag, 2019 #3539}. The use of generative AI in assessment design can facilitate plagiarism, as students can easily use AI-generated content to cheat on assignments {Crawford, 2023 #3540}. This issue poses a significant challenge to educators, who must ensure that assessment tasks are designed to promote academic integrity while also taking advantage of the benefits of the generative AI {Crawford, 2023 #3540}.

This paper addresses the challenges associated with the integration of generative AI in assessment design, specifically in the context of take-home assignments in software-related courses {Gilson, 2023 #3541}. The researchers propose a framework that provides guidelines for designing assessment tasks that incorporate generative AI, while also promoting academic integrity {Perkins, 2023 #3542}. The proposed framework offers a practical and flexible approach for educators seeking to integrate generative AI into their assessment design process {Fergus, 2023 #3533}.

The paper is organized as follows: Section 2 provides a brief overview of the related work in the area of generative AI and assessment design. Section 3 describes our research approach for designing take-home assignments that incorporate generative AI. Section 4 presents findings that demonstrate the application of the proposed framework in the context of software-related courses. Finally, Section 5 concludes the paper and provides directions for future research in the area of generative AI and assessment design.

1. Related Research
   * 1. Bloom taxonomy

Bloom's Taxonomy is a well-known framework that has been used for several decades to describe the various levels of cognitive skills required for learning. Developed by Benjamin Bloom in the 1950s, the taxonomy identifies six levels of cognitive skills, which are arranged in a hierarchy, starting from the lowest level of recall or memory to the highest level of creation or synthesis. The six levels of Bloom's Taxonomy include: remembering, understanding, applying, analyzing, evaluating, and creating. The taxonomy has become a popular tool for educators worldwide to design learning objectives, the curriculum, and assessments that focus on developing students' cognitive skills {Forehand, 2010 #1405}. It has been widely used in various fields, including education, psychology, and business, to design effective instructional strategies that enhance learning outcomes.

The framework has undergone several revisions over the years, with the most recent update released in 2001 by a team of educational psychologists led by Lorin Anderson. The revised taxonomy includes updated descriptions of each cognitive skill level, reflecting the current understanding of learning and teaching. Table 1 presents the comparison of generic questions of Bloom’s Taxonomy and Computer-Related Topics Questions of Bloom’s Taxonomy. Both share the same cognitive levels of Remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. However, the content and context of the questions differ between the two sets. The questions in generic questions are general in nature and can be applied to a wide range of subjects or topics while computer-related topics questions focus on concepts, principles, and skills within the field of computer science and technology.

Table 1: Comparison of Generic Questions and Computer-Related Topics Questions

|  |
| --- |
| Computer-related topics questions of Bloom Taxonomy |
| 1. Remembering:    * What is the difference between a compiler and an interpreter?    * What is the syntax for creating a new object in Java?    * What is the purpose of a firewall in network security? 2. Understanding:    * How do programming languages like Java and Python differ in their approach to memory management?    * What is the role of an operating system in managing computer resources?    * How does the client-server model work in web development? 3. Applying:    * Can you write a program in Python that takes user input and performs a specific task?    * How would you design a database schema to store information about users in a web application?    * Can you troubleshoot and solve a common networking issue? 4. Analyzing:    * How would you identify and fix a performance bottleneck in a software application?    * How would you analyze the security risks associated with a particular system or application?    * Can you break down the steps involved in designing and implementing a complex algorithm? 5. Evaluation:    * How would you assess the effectiveness of a particular software development methodology?    * Can you critique the design of a user interface and suggest improvement?    * How would you evaluate the ethical implications of using artificial intelligence in a particular application? 6. Creating:    * Can you design and implement a new software application that solves a real-world problem?    * How would you develop a new feature or functionality for an existing software product?    * Can you create a prototype of a new technology that has not yet been developed? |

* + 1. Assessment design

Assessment design is a critical aspect of higher education, as it provides important information on student learning and informs ongoing improvement efforts {Yeadon, 2023 #1720;Fergus, 2023 #1723}. One theme that emerges in the literature is the importance of aligning assessments with learning outcomes. {Chickering, 1987 #3543} argue that assessment should be closely tied to course objectives, providing evidence that students are meeting the intended goals. This requires careful attention to the development of clear and specific learning outcomes, as well as the design of assessments that measure those outcomes {Perkins, 2023 #1737}.

Another The use of generative AI in education has brought about various challenges, especially in the area of assessment design, particularly concerning academic integrity. Plagiarism is a significant concern as students can easily use AI-generated content to cheat in assignments, compromising the credibility and reliability of the assessment results. This paper addresses this issue by proposing a framework for designing take-home assignments in software-related courses incorporating generative AI and promoting academic integrity. The framework provides guidelines for designing assignments that prevent plagiarism and ensure that the assessment tasks align with the course learning objectives. It offers a practical and flexible approach to integrating generative AI into assessments while mitigating the risks associated with academic misconduct. consideration in assessment design is the use of rubrics. Many scholars emphasize the importance of rubrics in ensuring consistency and fairness in grading, as well as providing valuable feedback to students. {Brookhart, 2013 #3544} notes that rubrics can help students understand the expectations for their work and can encourage deeper engagement with the material. The literature also highlights the importance of incorporating a variety of assessment types in higher education {Mellar, 2018 #1756}. This includes both formative assessments, designed to provide ongoing feedback and support to students, and summative assessments, used to evaluate learning at the end of a course or program. {Brookhart, 2013 #3544@@author-year} notes that different assessment types may be more appropriate for different learning outcomes and that instructors should carefully consider which types will provide the most valid and reliable data.

* + 1. Bloom taxonomy and assessment design

Bloom's taxonomy is a widely used framework for designing and evaluating learning outcomes and assessments in higher education. The framework, developed by Benjamin Bloom in the 1950s, outlines a hierarchical structure of cognitive skills, ranging from lower-order thinking skills (remembering and understanding) to higher-order thinking skills (analyzing, evaluating, and creating). One theme that emerges in the literature is the importance of aligning assessments with the appropriate level of Bloom's taxonomy. This requires careful attention to the intended learning outcomes and the cognitive demands of the assessment {Wilson, 2016 #3547}. Anderson and Krathwohl (2001) suggest that assessments should be designed to measure the highest level of cognitive skills that students are expected to achieve, while also incorporating lower-level skills as necessary.

Another The use of generative AI in education has brought about various challenges, especially in the area of assessment design, particularly concerning academic integrity. Plagiarism is a significant concern as students can easily use AI-generated content to cheat in assignments, compromising the credibility and reliability of the assessment results. This paper addresses this issue by proposing a framework for designing take-home assignments in software-related courses incorporating generative AI and promoting academic integrity. The framework provides guidelines for designing assignments that prevent plagiarism and ensure that the assessment tasks align with the course learning objectives. It offers a practical and flexible approach to integrating generative AI into assessments while mitigating the risks associated with academic misconduct. consideration in using Bloom's taxonomy in assessment design is the importance of providing clear and specific learning objectives. This requires careful attention to the development of learning outcomes that are aligned with Bloom's taxonomy and that provide a clear roadmap for both instruction and assessment. Gronlund and Waugh (2009) emphasize the importance of developing objectives that are measurable and that specify the desired level of cognitive skill.

The literature also highlights the importance of using a variety of assessment types to measure different levels of Bloom's taxonomy. This includes both formative assessments, designed to provide ongoing feedback and support to students, and summative assessments, used to evaluate learning at the end of a course or program. Krathwohl (2002) notes that different assessment types may be more appropriate for different levels of cognitive skill and suggests that instructors should carefully consider which types will provide the most valid and reliable data.

* + 1. Generative AI and assessment design

How to design good assessment that could accurately measure students’ performance has been a holy grail in the educational research {Villarroel, 2018 #1748;Mellar, 2018 #1756;Pittaway, 2009 #1751;Van Den Berg, 2006 #1753}. Problems that affect academic integrity include students plagiarizing works without proper citation and seeking improper support through contract cheating, which has become a pervasive issue that requires urgent solutions {Bretag, 2019 #1747;Bretag, 2019 #1759;Angus, 2009 #1754}. The invention of generative AI tools such as ChatGPT has introduced a new challenge for educators. They now have to not only address human contract cheating but also tackle machine-based cheating, which has the potential to give all students the same high distinction marks {Fergus, 2023 #1723;Geerling, 2023 #1731;Ivanov, 2023 #1732;Lee, 2023 #1733}. It is crucial to have an assessment design framework that can guide educators in higher education to develop effective assessments. Such a framework is needed to counteract the impact of generative AI tools, ensuring the maintenance of teaching quality and fairness in the student performance evaluation {Kooli, 2023 #1725;Benuyenah, 2023 #1726;Ivanov, 2023 #1732;Gilson, 2023 #1738}. Additionally, the framework should aim to provide students with an optimal learning pathway while leveraging the benefits of generative AI tools {Dwivedi, 2023 #893;Yeadon, 2023 #1720;Gregorcic, 2023 #1721}.

1. Research Approach

The approach that the researchers used in this study is experimental research which is a widely used research method in various scientific fields, such as education, psychology, biology, and engineering. The main objective of experimental research is to examine causal relationships between variables through carefully designed experiments {Cook, 2002 #3548}. Experimental research is characterized by the manipulation of an independent variable (IV) while keeping all other variables constant. The dependent variable (DV) is the outcome that is measured to determine the effect of the IV. In addition to the IV and DV, experimental research often includes a control group that does not receive the experimental treatment {Trochim, 2001 #3549}. This helps to ensure that any changes observed in the experimental group are due to the treatment and not to other factors.

The researchers chose to use the experimental research approach as one of its The use of generative AI in education has brought about various challenges, especially in the area of assessment design, particularly concerning academic integrity. Plagiarism is a significant concern as students can easily use AI-generated content to cheat in assignments, compromising the credibility and reliability of the assessment results. This paper addresses this issue by proposing a framework for designing take-home assignments in software-related courses incorporating generative AI and promoting academic integrity. The framework provides guidelines for designing assignments that prevent plagiarism and ensure that the assessment tasks align with the course learning objectives. It offers a practical and flexible approach to integrating generative AI into assessments while mitigating the risks associated with academic misconduct. strengths is the ability to establish cause-and-effect relationships between variables. This is because experimental designs allow researchers to manipulate the independent variable and control for other variables that might affect the outcome. The high level of control also allows for the replication of the study and the ability to test the results across different settings and populations.

The research approach can be summarized in two steps. In the first step, Bloom's taxonomy was applied to classify questions from sample assignments. In the second step, feed these questions into ChatGPT and compare them with the marking rubric. Then use the rubric to grade the answers by ChatGPT.



Figure 1. Assessment Testing with ChatGPT

To conduct the study, the researchers selected take-home assessments from three courses in the Bachelor of Information Technology (BIT) and Bachelor of Software Engineering (BSE) programs at a university in Vietnam (now referred to as MUni). Table 2 shows the three courses’ spreads across programs and years. They are typical courses in the BIT and BSE not only at MUni but also in other universities around the world.

Table 2: Courses selected for the study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Course name** | **Programs** | **Year** | **Topics** |
| Advanced Programming Techniques | BSE | 3 | C++ |
| Enterprise Application Development | BSE | 2 | HTML, Javascript |
| Object Oriented Programming | BIT | 1 | Java |

Table 3 summarizes the duration and weight of the assessments. Within the three courses, researchers selected only take-home assessments for the study. The reason for this choice is that the researchers believe with invigilated tests, students do not have a chance to access any online resources plus AI generative tools like ChatGPT. Therefore, they can’t leverage the tool to support their answers. The choice of assessments also varies in duration. For example, some tests take a few days but for some tests, students have only 2-3 hours to complete.

Table 3: Assessment Selected for the Study.

|  |  |  |  |
| --- | --- | --- | --- |
| **Assessment name** | **Courses** | **Duration** | **Weight** |
| Group assessment | Advanced Programming Techniques | 2 weeks | 35% |
| Assessment 1 | Enterprise Application Development | 3 days | 30% |
| Final test | Object Oriented Programming | 2 hours | 40% |

1. Findings and discussions
   1. List of sample assessments and Bloom’s taxonomy

All the assessments demanded students to build a fictional application for a specific requirement such as delivery service, home sharing, or quiz management. Although the requirement is close to real life, all complex features have been removed to ensure students can deliver in a fixed timeframe (see Table 4). That’s why the Bloom level is just between 3 and 4.

Table 4: Assessment and Bloom's Taxonomy

|  |  |  |
| --- | --- | --- |
| **Assessment name** | **Questions** | **Bloom’s taxonomy** |
| Group assessment | Build a command line interface application for vacation home sharing | 3 (Applying) |
| Assessment 1 | Build a quiz management system | 4 (Analyzing), 6 (Creating) |
| Final test | Build Java command line interface application to manage delivery service | 3 (Applying) |

* 1. Marking answers from ChatGPT

Technically, ChatGPT has a word limit per response. It has been said that only 2048 characters are returned per response. It means that in any circumstance, students must break down the question and receive pieces of the answers and assemble them into complete work. The researchers’ experiment also followed this approach: break down the question and recombine the answers. Sometimes, questions can be fed directly without breaking down.

For example, the following code is the answer for the banking app from ChatGPT with a direct question:

class Account {

    private String accountNumber;

    private double balance;

    public Account(String accountNumber) {

        this.accountNumber = accountNumber;

        this.balance = 0;

    }

    public void credit(double amount) {

        balance += amount;

    }

    public void debit(double amount) {

        if (balance >= amount) {

            balance -= amount;

        } else {

            System.out.println("Insufficient balance.");

        }

    }

    public void transferTo(Account anotherAccount, double amount) {

        if (balance >= amount) {

            debit(amount);

            anotherAccount.credit(amount);

        } else {

            System.out.println("Insufficient balance.");

        }

    }

The following is the answer for question building quiz system:

<!DOCTYPE html>

<html>

<head>

<title>Quiz Maker</title>

<link rel="stylesheet" href="styles.css">

</head>

<body>

<h1>Quiz Maker</h1>

<div id="quiz-edit-container">

<h2>Quiz Editor</h2>

<label for="quiz-title">Quiz Title:</label>

<input type="text" id="quiz-title">

<button id="add-question-btn">Add Question</button>

<button id="save-quiz-btn">Save Quiz</button>

<div id="questions-container">

<!-- questions will be dynamically added here -->

</div>

</div>

After ChatGPT generated answers, the researchers invited lecturers teaching the courses and marking students to check with the rubric. The result is summarized in Table 5.

Table 5: Marking answers from ChatGPT

|  |  |
| --- | --- |
| **Assessment name** | **Probable marks** |
| Group assessment | 70% |
| Assessment 1 | 65% |
| Final test | 75% |

* 1. Towards a framework for take-home assessment design

Based on the results, the researcher generalizes the findings to construct a framework to guide take-home assessment so that students couldn’t exploit ChatGPT for their study and pass a course without substantial knowledge and skill required by the course. The list of rules is as follows:

* Use questions that are at a level higher than 3 and 4, with a recommendation for levels 5 and 6. Our findings show that most of the questions in the 3 tests are in levels 3-4 of Bloom’s taxonomy. This gives some difficulties to ChatGPT to provide exact answers {Yeadon, 2023 #1720;Emenike, 2023 #1722;Fergus, 2023 #1723}.
* Use complex questions that do not allow students to receive all the answers from ChatGPT at once and submit them. Assembling and recompiling them would be excellent for learning and are desired skills for software engineers {Geerling, 2023 #1731;Lee, 2023 #1733;Gilson, 2023 #1738}. The lecturer who gave the Final test leveraged ChatGPT to revise questions in a more complex way that students could not easily get straight answers from ChatGPT.
* Include images in questions, although this can be challenging for ChatGPT, which is primarily based on text. While ChatGPT can generate diagrams, feeding images into the system is not currently supported. For example, assessment 1 included many images that could prevent students from copying and pasting into ChatGPT. This could be a temporal solution as OpenAI could release new version with image recognition support and students can combine image related AI for producing results.
* Leverage invigilated in-class or online tests with camera recording to neutralize ChatGPT usage. Use long questions with many details that need to be combined to get an answer and require students to complete them within a short time, such as 1-2 hours. This could be a quick fix for AI related plagiarism that is very hard to detect.

The proposed assessment design framework is summarized in Figure 2.

A picture containing text, screenshot, font

Description automatically generated

Figure 2. Assessment design framework

1. Conclusion

In conclusion, the use of generative AI in education has revolutionized the way assessments are designed and conducted, bringing about various benefits and challenges {Gašević, 2023 #1739;Crawford, 2023 #3}. One of the primary concerns associated with this technology is academic integrity, as students may use AI-generated content to cheat in assignments, compromising the reliability and credibility of assessment results {Lo, 2023 #1724;Benuyenah, 2023 #1726}. However, the framework proposed in this paper offers a practical and flexible approach to incorporating generative AI in software-related courses' take-home assignments while promoting academic integrity. By providing guidelines that prevent plagiarism and align the assessment tasks with the course learning objectives, this framework helps to mitigate the risks associated with academic misconduct while enhancing the educational experience. The study has some limitations in its design, for example, validation testing of the framework was not done with breath and depth.

Future research could extend to other courses in various disciplines, including electronics, mechanics, food technology, and more. This would allow the framework to be tested in different contexts. It is expected that additional rules or modified rules will be developed to suit these new disciplines.

1. References